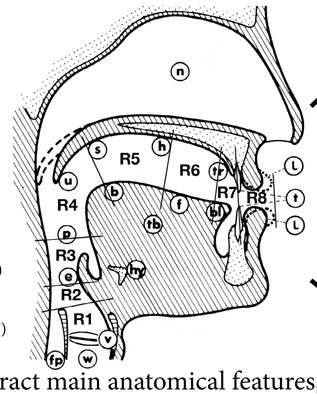
The GNU project "gnuspeech": a real-time articulatory synthesis system for spoken English, and a tool for articulatory phonetics research

by David R. Hill, University of Calgary

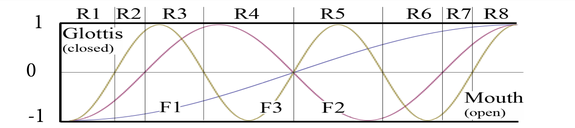
Background to the work ①

Fant & Pauli (1974): a sensitivity analysis of human vocal tract resonances (formants), showed their frequencies depend on constrictions affecting the balance between kinetic (1) and potential (-1) energy in the standing waves.

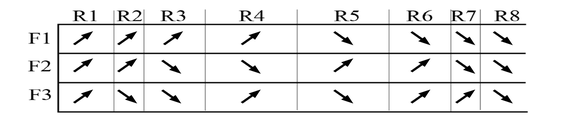
KEY
b - back of tongue
bl - tongue blade
e - epiglottis
f - front of tongue
fp - food passage (esophagus)
h - hard palate
hy - hyoid bone
l - lips (plus protrusion)
n - nasal passage
p - pharynx
t - teeth
tb - tongue body
tr - teeth ridge (alveolar ridge)
s - soft palate & velum
u - uvula/soft palate tip/velum
v - vocal folds (vocal "chords")
w - windpipe (trachea)



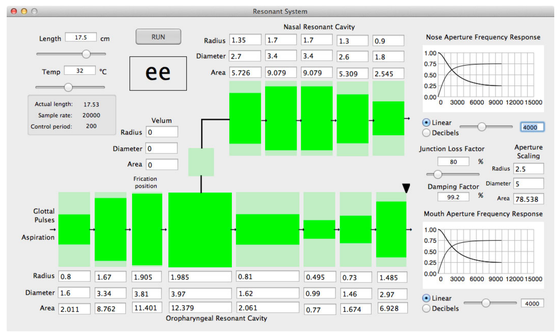
The human vocal tract main anatomical features, with the Distinctive Region Model divisions indicated



DRM regions related to formant sensitivity functions



Effect on formants of constriction in the DRM regions



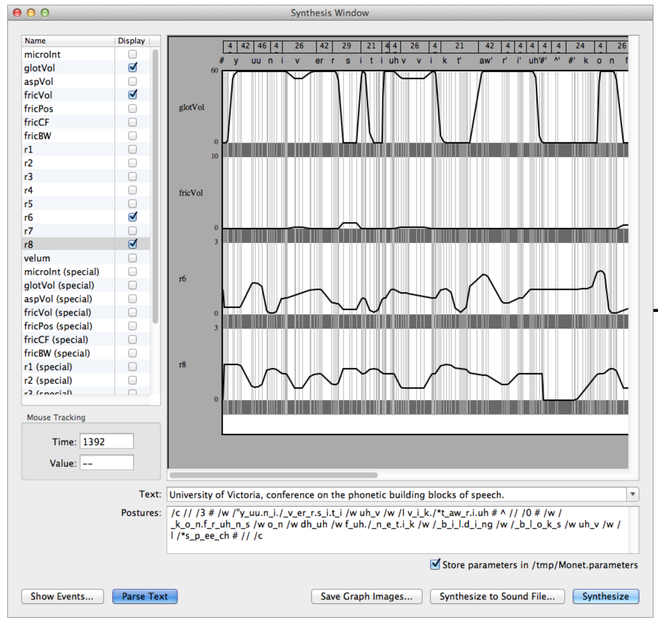
The "TRAcT" main window
(illustrates the waveguide "tube" model)

Implementation tools

Monet is the *gnuspeech* database editor. It allows articulatory posture and dynamic transition data to be created and linked, by rules, to phonetic symbols, with specified "event-based" asynchronous timing. Monet can also apply modifiable intonation and rhythm models.

Monet includes a number of sub-modules and displays for these purposes. The synthesis window is shown at the right, and part of the rule window below.

TRAcT provides a direct interface to all the waveguide tube model controls. It is used in the experimental derivation of posture and transition data during database development for a new language, using Monet together with listening trials and spectrographic analysis.



The "Monet" synthesis window: provides controls, input, parsed text and selected parameters, as well as access to intonation control, display and modification

Overview

A speech synthesis system was implemented on a NeXT computer, using an acoustic tube model to emulate the human vocal tract and nasal cavities. Supporting software allows the creation/modification of the necessary databases needed for dynamic "event-based" control of the acoustic tube as required to synthesise spoken English (or other languages) and facilitate phonetic research.

Text-to-phone conversion software , including effective models of rhythm and intonation, provides real-time computer conversion of unrestricted English text to spoken output based on pronouncing dictionaries. The system has been used for phonetics research on stop-bursts.

Subjective testing reveals significant user preference for this speech compared to other systems.

The system is currently being updated and ported to both the Macintosh OS X, and GNU/Linux using GNUStep. Those able to contribute are most welcome to join the project team at:

savannah.gnu.org/projects/gnuspeech
www.gnu.org/software/gnuspeech/

Participation is strongly encouraged

Origin of the tube model ②

Carré and his co-workers used Fant & Pauli's work as a basis for their "Distinctive Region Model" (DRM), which they said was capable of "producing optimal formant frequency modulation" for all the formants. The regions bear a noticeable relationship to the human articulator locations.

The basis of gnuspeech ③

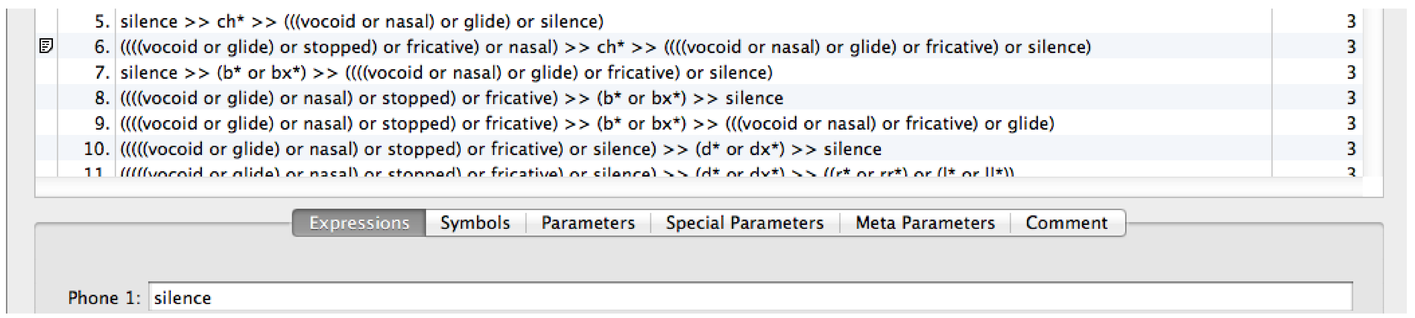
gnuspeech is based on the DRM concept, using the tube model that emulates the human vocal tract directly. Using the DRM means only 8 sections are needed, the computation is reduced, and the control data is related directly to articulation.

Immediate Future work ④

(1) The database creation/modification components of "Monet"—currently the most important element missing from the port—need to be ported.

(2) Meta-parameters more easily linked to articulation—jaw rotation, tongue height, etc.,— should be introduced.

(3) Other modules need porting: (a) "PrEditor", which allows User and Application dictionaries to be created to supplement the Main dictionary to allow special or unusual pronunciations; (b) "BigMouth" which provides a speech service; and certain utilities such as "Speech Manager" to vary the system-level environment; and (c) "ServerTest" that provides a direct interface to the TextToSpeech server. (4) A standalone Speech Server that can run directly under GNU/Linux, independent of GNUStep, is also needed.



The "Monet" rule manager selects transitions and timing for n-phone combinations to be created, edited, stored & applied

Acknowledgement of those involved in this long term project is provided on the GNU project home page: www.gnu.org/software/gnuspeech/ (accessed 2014-06-17)
Contact: hilld@ucalgary.ca ● Phone: 604-947-9362 ● More information, including audio files of the speech output, is available on the author's website: pages.cpsc.ucalgary.ca/~hill (accessed 2014-06-15)
Sponsorship by FireThorne® Consulting is gratefully acknowledged (www.firethorne.com—accessed 2014-06-17)